SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 882

January 1990

ANTICIPATED BEHAVIOR OF LONG-STEP ALGORITHMS FOR LINEAR PROGRAMMING

By

Shinji Mizuno*
Michael J. Todd†
Yinyu Ye

*Research was supported in part by Grant-in-Aid 63490010 for General Scientific Research of the Ministry of Education, Science and Culture, Japan.

†Research was supported in part by NSF grant DMS-8904406 and ONR contract N-00014-87-K0212.

Anticipated behavior of long-step algorithms for linear programming

Shinji MIZUNO *

Department of Industrial Engineering and Management,

Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152, Japan

Michael J. TODD †

School of Operations Research and Industrial Engineering,

Cornell University, Ithaca, New York 14853, USA.

Yinyu YE

Department of Management Sciences,

The University of Iowa, Iowa City, Iowa 52242, USA.

December 1989

Abstract: We provide a probabilistic analysis of the second order term that arises in path-following algorithms for linear programming. We use this result to show that two such methods, algorithms generating a sequence of points in a neighborhood of the central path and in its relaxation, require a worst-case number of iterations that is $O(nL)$ and an anticipated number of iterations that is $O(\log(n)L)$. The second neighborhood spreads almost all over the feasible region so that the generated points are close to the boundary rather than the central path. We also

*Research was supported in part by Grant-in-Aid 63490010 for General Scientific Research of the Ministry of Education, Science and Culture, Japan.

†Research was supported in part by NSF grant DMS-8904406 and ONR contract N-00014-87-K0212.

propose a potential reduction algorithm which requires the same order of number of iterations as the path-following algorithms.

Key words: Linear Programming, interior point algorithms, path-following algorithms, potential reduction algorithms.

1 Introduction

We consider linear programs in the following standard form:

$$\begin{aligned} (P) \quad & \text{minimize} \quad c^T x \\ & \text{subject to} \quad Ax = b,\ x \geq 0, \end{aligned}$$

where $c \in R^n$, $A \in R^{m \times n}$ and $b \in R^m$ are given, $x \in R^n$, and ^{T} denotes transpose. The dual to (P) can be written as

$$\begin{aligned} (D) \quad & \text{maximize} \quad b^T y \\ & \text{subject to} \quad A^T y + s = c,\ s \geq 0, \end{aligned}$$

where $y \in R^m$ and $s \in R^n$. The components of s are called dual slacks. Denote by $\mathcal{F}$ the set of all x and s that are feasible for the primal and dual, respectively. We also denote the set of points $(x, s) > 0$ in $\mathcal{F}$ by $\mathcal{F}^0$. Assume that $\mathcal{F}^0 \neq \phi$.

Megiddo [11], Sonnevend [18], and Bayer and Lagarias [2] analyzed the central path which is expressed as

$$\mathcal{C} = \left\{ (x, s) \in \mathcal{F}^0 : Xs = \frac{x^T s}{n} e \right\}$$

in primal-dual form, where e denotes the vector of ones and $X = \text{diag}(x)$ denotes the diagonal matrix with diagonal entries equal to the elements of x. Renegar [16], Gonzaga [4], Vaidya [20], Kojima, Mizuno and Yoshise [7,8], and Monteiro and Adler [14,15] proposed algorithms that generate a sequence of points in a neighborhood of the central path $\mathcal{C}$ and move in a direction that tries to find a new central point. These are called central path-following methods. The original primal-dual path-following algorithm presented by Kojima, Mizuno and Yoshise [7] uses the neighborhood

$$\left\{ (x, s) \in \mathcal{F}^0 : Xs \geq 0.5 \frac{x^T s}{n} e \right\}.$$

The algorithm requires at most $O(nL)$ iterations. Then Kojima, Mizuno, and Yoshise [8] and Monteiro and Adler [14,15] modified the algorithm by using the neighborhood

$$\mathcal{N}(\beta) = \left\{ (x, s) \in \mathcal{F}^0 : \|Xs - \mu e\| \leq \beta \mu \quad \text{where} \quad \mu = \frac{x^T s}{n} \right\}$$

for some $\beta \in (0,1)$ so that it requires at most $O(n^{0.5}L)$ iterations theoretically. Here $\|\cdot\|$ without subscript designates the l_2-norm.

Recently, Mizuno, Todd and Ye [12] analyzed an anticipated $O(n^{1/4}L)$-iteration path-following algorithm which generates a sequence of points moving inside of $\mathcal{N}(\beta)$ for $\beta = 0.5$. The notion of "anticipated" behavior was also considered in Gonzaga and Todd [5]. The idea is that, at each iteration, we make an unrigorous but plausible assumption on the probability distribution of the data of the problem, and then consider the behavior which occurs with high probability (converging to 1 as $n \to \infty$) at that iteration. The anticipated number of iterations is then defined to be the number of iterations required if this high probability behavior actually occurs at least once every ten, say, iterations.

In this paper, we show that, for each $\beta \in (0,1)$, a central path-following algorithm generating a sequence of points in either

$$\mathcal{N}_\infty(\beta) = \left\{(x,s) \in \mathcal{F}^0 : \|Xs - \mu e\|_\infty \le \beta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n}\right\}$$

or

$$\mathcal{N}_\infty^-(\beta) = \left\{(x,s) \in \mathcal{F}^0 : \|Xs - \mu e\|_{-\infty} \le \beta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n}\right\}$$

requires an anticipated number of iterations that is $O(\log(n)L)$ and a worst-case number of iterations that is $O(nL)$. Here $\|\cdot\|_\infty$ designates the ℓ_∞-norm and

$$\|z\|_{-\infty} = -\min\{z_1, z_2, \ldots, z_n, 0\} \text{ for each } z \in R^n.$$

Note that $\|\cdot\|_{-\infty}$ is not a norm. The neighborhood used by Kojima, Mizuno, and Yoshise [7] equals $\mathcal{N}_\infty^-(\beta)$ for $\beta = 0.5$ and the worst-case number of iterations in our algorithm is of the same order as theirs. We easily see that

$$\mathcal{C} \subset \mathcal{N}(\beta) \subset \mathcal{N}_\infty(\beta) \subset \mathcal{N}_\infty^-(\beta) \subset \mathcal{F}^0 \text{ for each } \beta \in (0,1).$$

Our result indicates that when we use a wider neighborhood of the central path, although the worst-case number of iterations grows, the anticipated number of iterations falls.

An interesting result of our algorithm is that both the worst-case and anticipated numbers of iterations are obtained for each constant $\beta \in (0,1)$ independent of n. So the

algorithm can generate a sequence of points in a large area of the feasible region. Note that we have

$$\mathcal{N}_\infty^-(1) = \mathcal{F}^0.$$

So when β is close to 1, the set $\mathcal{N}_\infty^-(\beta)$ spreads almost all over the feasible region $\mathcal{F}$. Hence the points generated at each iteration of the algorithm are close to the boundary rather than the central path. Moreover the search direction of our algorithm approaches that of the original primal-dual affine scaling algorithm (Monteiro, Adler, and Resende [13]) if β goes to 1.

We also describe a potential reduction algorithm. Although the search direction of the potential reduction algorithm is the same as that of the path-following algorithm, the step size is determined as the minimum point of the potential function in the neighborhood $\mathcal{N}_\infty^-(\beta)$. The primal-dual potential function introduced by Todd and Ye [19] is

$$\psi(x,s) = \rho \log(x^T s) - \sum_{j=1}^{n} \log(x_j s_j), \tag{1}$$

where $\rho > n$. Using this, Ye [21], Freund [3], Anstreicher and Bosch [1], and Kojima, Mizuno and Yoshise [6] have developed $O(n^{0.5}L)$-iteration potential reduction algorithms with the choice of $\rho = n + \theta(n^{0.5})$. However, practical experiments indicate that a big ρ is much better (McShane et al. [10] and Lustig et al. [9]). We show that if we set $\rho = n + \theta(n^2)$, the potential reduction algorithm requires an anticipated number of iterations that is $O(\log(n)L)$ and a worst-case number of iterations that is $O(nL)$.

2 A primal-dual path-following algorithm

Let $\mathcal{N}$ be a subset of R^{2n} such that

$$\mathcal{C} \subset \mathcal{N} \subset \mathcal{F}^0.$$

Assume that we have initial feasible points x^0 of (P) and (y^0, s^0) of (D) such that $(x^0, s^0) \in \mathcal{N}$ and $(x^0)^T s^0 \leq 2^L$, where L denotes the size of (P). Let γ be a constant in $(0,1)$. Given

a pair $(x^k, s^k) \in \mathcal{N}$ and $\mu^k = (x^k)^T s^k/n$, we generate the search direction using the primal-dual scaling method (Kojima et al. [7])

$$\begin{aligned} S^k \Delta x + X^k \Delta s &= \gamma \mu^k e - X^k s^k, \\ A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0. \end{aligned} \tag{2}$$

Let

$$\begin{aligned} x(\theta) &= x^k + \theta \Delta x, \\ y(\theta) &= y^k + \theta \Delta y, \\ s(\theta) &= s^k + \theta \Delta s. \end{aligned}$$

We determine the step parameter

$$\bar{\theta} = \max \{\theta : (x(\theta), s(\theta)) \in \mathcal{N}\}. \tag{3}$$

Then compute the next points $x^{k+1} = x(\bar{\theta})$, $y^{k+1} = y(\bar{\theta})$, and $s^{k+1} = s(\bar{\theta})$.

Thus we propose the following path-following algorithm based on γ and $\mathcal{N}$.

Algorithm 1 *Given* $(x^0, s^0) \in \mathcal{N}$ *and* $(x^0)^T s^0 \leq 2^L$; *set* $k = 0$;

while $(x^k)^T s^k \geq 2^{-L}$ **do**

begin

compute Δx, Δy, *and* Δs *using (2);*

compute $\bar{\theta}$ *using (3);*

set $x^{k+1} = x(\bar{\theta})$, $y^{k+1} = y(\bar{\theta})$, *and* $s^{k+1} = s(\bar{\theta})$;

k=k+1;

end.

Note that, if $\gamma = 0$, the direction of this algorithm equals that of the original primal-dual affine scaling algorithm. So we can set the direction almost equal to that of the affine scaling algorithm by choosing γ close to 0.

The neighborhoods $\mathcal{N}(\beta)$, $\mathcal{N}_\infty(\beta)$, and $\mathcal{N}_\infty^-(\beta)$ presented in the introduction are defined by restricting the size of $Xs - \mu e$, which expresses a deviation from a center. Hence, in order to see the size of parameter $\bar{\theta}$, we evaluate the size of $X(\theta)s(\theta) - \mu(\theta)e$ where $X(\theta) = \mathrm{diag}(x(\theta))$ and $\mu(\theta) = x(\theta)^T s(\theta)/n$. We easily see that

$$\mu(\theta) = (1-\theta)\mu^k + \theta\gamma\mu^k, \tag{4}$$

$$X(\theta)s(\theta) - \mu(\theta)e = (1-\theta)(X^k s^k - \mu^k e) + \theta^2 \Delta X \Delta s, \tag{5}$$

where $\Delta X = \mathrm{diag}(\Delta x)$. So we want to know the size of $\Delta X \Delta s$. Note that $\Delta X \Delta s$ equals the error $X(\theta)s(\theta) - \mu(\theta)e$ when we apply Newton's method, i.e., $\theta = 1$. Let

$$\begin{aligned} p &= (X^k)^{-0.5}(S^k)^{0.5}\Delta x, \\ q &= (X^k)^{0.5}(S^k)^{-0.5}\Delta s, \\ r &= (X^k S^k)^{-0.5}(\gamma\mu^k e - X^k s^k), \end{aligned} \tag{6}$$

and $U = \{z : AX^{0.5}S^{-0.5}z = 0\}$. Then the system (2) can be written as

$$p + q = r,$$

$$p \in U \text{ and } q \in U^\perp,$$

where $U^\perp$ denotes the complementary subspace of U. Thus p is the projection of r on U, and q the projection of r on $U^\perp$. Since $\Delta X \Delta s = Pq$, we analyze Pq in the next section, where $P = \mathrm{diag}(p)$.

3 Analysis of Pq

Throughout this section, all symbols except for p, q, r, U, and P $(= \mathrm{diag}(p))$ have a local meaning (for example, Q does not mean $\mathrm{diag}(q)$). For an $r \in R^n$ and a subspace U of R^n, let p and q be the projection of r on U and $U^\perp$ respectively. Then the analysis of the vector Pq is useful to evaluate the step size at each iteration of Algorithm 1. At first we give a deterministic analysis of Pq and then a probabilistic analysis.

Lemma 1 *Let r be a vector in R^n and U be a subspace of R^n. Let p and q be the projection of r on U and $U^\perp$ respectively. Then*

$$-\frac{\|r\|^2}{4} \leq p_j q_j \leq \frac{r_j^2}{4}.$$

Proof. The right inequality easily follows from $p_j + q_j = r_j$ for each j. Then the left inequality follows from

$$\begin{aligned} p_j q_j &\geq \sum_{p_i q_i < 0} p_i q_i \\ &= -\sum_{p_i q_i \geq 0} p_i q_i \quad (\text{since } p^T q = 0) \\ &\geq -\sum_{p_i q_i \geq 0} \frac{r_i^2}{4} \\ &\geq -\frac{\|r\|^2}{4}. \end{aligned}$$

□

The above lemma implies that

$$\begin{aligned} \|Pq\|_{-\infty} &\leq \|r\|^2/4 \leq n\|r\|_\infty^2/4, \\ \|Pq\|_{+\infty} &\leq \|r\|_\infty^2/4, \end{aligned}$$

where $\|z\|_{+\infty} = \max\{z_1, z_2, \ldots, z_n, 0\}$ for each $z \in R^n$. Then $\|z\|_\infty = \max\{\|z\|_{-\infty}, \|z\|_{+\infty}\}$. These bounds cannot be improved by much in the worst case, since we have that $\|Pq\|_{-\infty} = (n-1)\|r\|_\infty^2/4$ and $\|Pq\|_{+\infty} = \|r\|_\infty^2/4$ when

$$\begin{aligned} r &= e = (1, 1, \ldots, 1)^T, \\ p &= (1/2, 1/2, \ldots, 1/2, (1+\sqrt{n})/2)^T, \\ q &= (1/2, 1/2, \ldots, 1/2, (1-\sqrt{n})/2)^T. \end{aligned}$$

Since the bound of $\|Pq\|_{-\infty}$ is much larger than that of $\|Pq\|_{+\infty}$, we shall get a better bound of $\|Pq\|_{-\infty}$ by a probabilistic analysis. In the theorem below, a random d-subspace means a d-dimensional subspace generated from a distribution that is invariant under orthogonal transformations. For instance, we can assume that it is the null space of a

random matrix of appropriate dimensions, each entry of which has a standard normal distribution.

Theorem 2 *Let $r \in R^n$ be fixed, and let U be a random d-subspace of R^n. Let p and q be the projection of r on U and $U^\perp$ respectively. Then*

$$\Pr\left(\|Pq\|_{-\infty} \leq 3(\log(n)/n)\|r\|^2\right) \to 1 \text{ as } n \to \infty.$$

We divide the proof of the theorem into two lemmas. Because p and q are homogeneous of degree 1 in $\|r\|$, we assume henceforth without loss of generality that r is scaled so that

$$t = r/2 \text{ satisfies } \|t\| = 1.$$

Let $Q = [t, Y]$ be an orthogonal $n \times n$ matrix. If we express the vector p using the basis Q, we get the following lemma.

Lemma 3 (Mizuno, Todd and Ye [12]) *We can write*

$$p = (1 + \mu)t + \nu Y z, \tag{7}$$

where $\mu^2 + \nu^2 = 1$ and z is uniformly distributed on the unit sphere in R^{n-1}.

From (7) and $p + q = 2t$, we have

$$q = (1 - \mu)t - \nu Y z,$$

so that

$$\begin{aligned} Pq &= \nu^2 t^2 - 2\mu\nu T Y z - \nu^2 (Yz)^2 \\ &= -(Yz)^2 + (\nu t - \mu Y z)^2 \\ &\geq -\|Yz\|_\infty^2 e, \end{aligned}$$

where $T = \text{diag}(t)$, and t^2, $(Yz)^2$, and $(\nu t - \mu Y z)^2$ denote the vectors whose components are the squares of those of t, Yz, and $\nu t - \mu Y z$ respectively. From the above inequality and the following lemma, we can easily prove Theorem 2.

Lemma 4 *Let $Q = [t, Y]$ be an orthogonal matrix. If z is uniformly distributed on the unit sphere in R^{n-1},*

$$\Pr\left(\|Yz\|_\infty \leq \sqrt{3\frac{\log n}{n}}\right) \to 1 \text{ as } n \to \infty.$$

Proof. Since z is uniformly distributed on the unit sphere in R^{n-1}, z can be generated as follows: $z = \lambda/\|\lambda\|$, where $\lambda \sim N(0, I)$ (the standard normal distribution in R^{n-1}). Now $\|\lambda\|^2$ is a χ^2 random variable with $n-1$ degrees freedom, so

$$\begin{aligned} \mathrm{E}(\|\lambda\|^2) &= n-1, \\ \mathrm{Var}(\|\lambda\|^2) &= 2(n-1). \end{aligned}$$

From Chebychev's inequality, we have

$$\Pr\left(\|\lambda\| \geq (1-\epsilon)\sqrt{n-1}\right) \to 1 \text{ as } n \to \infty \tag{8}$$

for any $\epsilon > 0$.

Let λ_0 be a standard normal variable, and let $\lambda' = (\lambda_0, \lambda)$, also $N(0, I)$ but in R^n. Then $\|\lambda'\|_\infty = \max\{\mu_j : j = 0, 1, 2, \ldots, n-1\}$ where $\mu_j = |\lambda_j|$ has the positive normal distribution. Then $1 - F(x) = 2(1 - N(x))$ where F is the distribution function of μ, and N is the normal distribution function. It now follows from results in extreme value theory (Resnick [17], pp. 42 and 71) that

$$\Pr\left(\|\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

Since $Q\lambda'$ is also $N(0, I)$,

$$\Pr\left(\|Q\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

Now we have

$$\|Y\lambda\|_\infty \leq \|Q\lambda'\|_\infty + \|\lambda_0 t\|_\infty.$$

Since $\|t\| = 1$,

$$\Pr\left(\|\lambda_0 t\|_\infty \leq \epsilon\sqrt{\log n}\right) \to 1 \text{ as } n \to \infty$$

for any $\epsilon > 0$. From the above relations and (8), we get the result of the lemma. □

4 A path-following algorithm based on the ℓ_∞-norm

Based on the deterministic and probabilistic analysis of Pq in the previous section, we develop a large-step path-following algorithm for linear programming in this section. The main result of this section is expressed as follows.

Theorem 5 *Let $\mathcal{N} = \mathcal{N}_\infty(\beta)$ and $\gamma = 1 - \beta$ for $\beta \in (0,1)$. Then Algorithm 1 requires an anticipated number of iterations that is $O(\log(n)L)$, with a worst-case bound of $O(nL)$.*

We shall prove the theorem at the end of this section. From (4) and $\gamma = 1 - \beta$, we see

$$\mu^{k+1} = \mu(\bar{\theta}) = (1 - \beta\bar{\theta})\mu^k, \tag{9}$$

where $\bar{\theta}$ is the step parameter defined by (3). We remark that $\bar{\theta}$ can be calculated by solving $2n$ single-variable quadratic equations. Here we bound the parameter $\bar{\theta}$ from below as follows.

Lemma 6 *Let $\beta \in (0,1)$, $\gamma \in (0,1)$, and $\mathcal{N} = \mathcal{N}_\infty(\beta)$. Let*

$$\theta_1 = \min\left\{1, \frac{\beta\gamma\mu^k}{\|\Delta X \Delta s\|_\infty}\right\}$$

and $\bar{\theta}$ be the step parameter at the kth iteration of Algorithm 1. Then $\bar{\theta} \geq \theta_1$.

Proof. From (4), (5), and the definition of $\mathcal{N}_\infty(\beta)$, for each $\theta \in [0, \theta_1]$

$$\begin{aligned} \|X(\theta)s(\theta) - \mu(\theta)e\|_\infty &\leq (1-\theta)\|X^k s^k - \mu^k e\|_\infty + \theta^2 \|\Delta X \Delta s\|_\infty \\ &\leq (1-\theta)\beta\mu^k + \theta\beta\gamma\mu^k \\ &= \beta\mu(\theta). \end{aligned}$$

Since $\beta < 1$, we see $X(\theta)s(\theta) > 0$ for each $\theta \in [0, \theta_1]$. From the continuity of $x(\theta)$ and $s(\theta)$ with respect to θ, $x(\theta) > 0$ and $s(\theta) > 0$. It is easy to see that $Ax(\theta) = b$ and $A^T y(\theta) + z(\theta) = c$. Hence we have $(x(\theta), s(\theta)) \in \mathcal{N}_\infty(\beta)$ for each $\theta \in [0, \theta_1]$ which implies $\bar{\theta} \geq \theta_1$. □

In order to see the size of θ_1, we need to estimate $\|\Delta X \Delta s\|_\infty$. As we have shown in the previous section, $\|\Delta X \Delta s\|_\infty$, which is equal to $\|Pq\|_\infty$, is measured by the size of r. So we evaluate the size of r.

Lemma 7 *Let $\beta \in (0,1)$ and $\gamma = 1-\beta$. If $(x^k, s^k) \in \mathcal{N}_\infty(\beta)$ then*

$$0 \geq r_j \geq -\frac{2\beta\sqrt{\mu^k}}{\sqrt{1+\beta}} \quad \textit{for each } j,$$

where r is defined by (6).

Proof. From (6) and $\gamma = 1-\beta$, each component of r is given by

$$r_j = \frac{(1-\beta)\mu^k}{\sqrt{x_i^k s_j^k}} - \sqrt{x_j^k s_j^k}.$$

Since $(x^k, s^k) \in \mathcal{N}_\infty(\beta)$, we have

$$\sqrt{(1-\beta)\mu^k} \leq \sqrt{x_j^k s_j^k} \leq \sqrt{(1+\beta)\mu^k} \quad \text{for each } j. \tag{10}$$

Obviously, r_j is a monotonically decreasing function in terms of $\sqrt{x_j^k s_j^k}$ in the interval of (10). Thus,

$$0 \geq r_j \geq -\frac{2\beta\sqrt{\mu^k}}{\sqrt{1+\beta}}.$$

□

An interesting observation from Lemma 7 is that at least one of Δx_j and Δs_j is less than zero for all j. Lemma 7 also indicates that

$$\|r\|_\infty \leq \frac{2\beta\sqrt{\mu^k}}{\sqrt{1+\beta}}.$$

Proof of Theorem 5. From the above inequality, Lemma 1, and $\Delta X \Delta s = Pq$,

$$\begin{aligned} \|\Delta X \Delta s\|_\infty &\leq \|r\|^2/4 \\ &\leq n(\|r\|_\infty)^2/4 \\ &\leq \frac{\beta^2 n \mu^k}{1+\beta}. \end{aligned}$$

From Lemma 6,

$$\bar{\theta} \geq \theta_1 \geq \min\left\{1, \frac{(1+\beta)\gamma}{\beta n}\right\}. \tag{11}$$

This inequality, together with (9), implies that

$$\mu^{k+1} \leq \left(1 - \min\left\{\beta, \frac{(1+\beta)\gamma}{n}\right\}\right)\mu^k.$$

Hence the number of iterations of Algorithm 1 is bounded by $O(nL)$.

Let $\alpha = 3$. Suppose that we have

$$\|\Delta X \Delta s\|_{-\infty} \leq \alpha(\log(n)/n)\|r\|^2 \tag{12}$$

at the kth iteration of Algorithm 1. Theorem 2 indicates that the above inequality holds with high probability when $U = \{z : AX^{0.5}S^{-0.5}z = 0\}$ is considered as a random subspace and n is sufficiently large. From Lemma 1, Lemma 7, and (12), we have

$$\begin{aligned}
\|\Delta X \Delta s\|_\infty &\leq \max\left\{\alpha(\log(n)/n)\|r\|^2, \|r\|_\infty^2/4\right\} \\
&\leq \alpha\log(n)\|r\|_\infty^2 \\
&\leq \frac{4\alpha\beta^2\log(n)\mu^k}{1+\beta}.
\end{aligned}$$

So

$$\bar{\theta} \geq \theta_1 \geq \min\left\{1, \frac{(1+\beta)\gamma}{4\alpha\beta\log(n)}\right\} \tag{13}$$

and

$$\mu^{k+1} \leq \left(1 - \min\left\{\beta, \frac{(1+\beta)\gamma}{4\alpha\log(n)}\right\}\right)\mu^k.$$

Hence Algorithm 1 will terminate if the inequality (12) actually holds at $O(\log(n)L)$ iterations. Therefore the anticipated number of iterations of Algorithm 1 is $O(\log(n)L)$.

□

5 A path-following algorithm in a wide region

In this section, we further relax the neighborhood $\mathcal{N}_\infty(\beta)$ to $\mathcal{N}_\infty^-(\beta)$. As in the original affine scaling algorithm, this condition mainly prevents the iterative points from going out of the interior of the feasible region.

The main result of this section is described as follows.

Theorem 8 *Let $\beta \in (0,1)$ and $\gamma \in (0,1)$ be constants such that $\gamma \leq 2(1-\beta)$. Let $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Then Algorithm 1 requires an anticipated number of iterations that is $O(\log(n)L)$, with a worst-case bound of $O(nL)$.*

We shall prove Theorem 8 from the following two lemmas.

Lemma 9 *Let $\beta \in (0,1)$, $\gamma \in (0,1)$, and $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Let*

$$\theta_2 = \min\left\{1, \frac{\beta\gamma\mu^k}{\|\Delta X \Delta s\|_{-\infty}}\right\}.$$

and $\bar{\theta}$ be the step parameter at kth iteration of Algorithm 1. Then $\bar{\theta} \geq \theta_2$.

Proof. From (4) and (5), for each $\theta \in [0, \theta_2]$

$$\begin{aligned} X(\theta)s(\theta) - \mu(\theta)e &= (1-\theta)(X^k s^k - \mu^k e) + \theta^2 \Delta X \Delta s \\ &\geq -\left((1-\theta)\|X^k s^k - \mu^k e\|_{-\infty} + \theta^2 \|\Delta X \Delta s\|_{-\infty}\right) e \\ &\geq -\left((1-\theta)\beta\mu^k - \theta\beta\gamma\mu^k\right) e \\ &= -\beta\mu(\theta)e. \end{aligned}$$

Hence, as in the proof of Lemma 7, we have $(x(\theta), s(\theta)) \in \mathcal{N}_\infty^-(\beta)$ for each $\theta \in [0, \theta_2]$ which implies $\bar{\theta} \geq \theta_2$. □

Lemma 10 *Let $\beta \in (0,1)$ and $\gamma \in (0,1)$ be constants such that $\gamma \leq 2(1-\beta)$. If $(x^k, s^k) \in \mathcal{N}_\infty^-(\beta)$ then*

$$\|r\|^2 \leq n\mu^k,$$

where r is defined by (6).

Proof. We see the result as follows:

$$\begin{aligned} \|r\|^2 &= \sum_{j=1}^n \frac{(\gamma\mu^k - x_j^k s_j^k)^2}{x_j^k s_j^k} \\ &= \sum_{j=1}^n \left(\frac{(\gamma\mu^k)^2}{x_j^k s_j^k} - 2\gamma\mu^k + x_j^k s_j^k\right) \\ &\leq \frac{n(\gamma\mu^k)^2}{(1-\beta)\mu^k} - 2n\gamma\mu^k + n\mu^k \quad (\text{since } x_j^k s_j^k - \mu^k \geq -\beta\mu^k) \\ &\leq n\mu^k \quad (\text{since } \gamma \leq 2(1-\beta)\ ). \end{aligned}$$

□

Proof of Theorem 8. From Lemma 1 and Lemma 10, we have

$$\|\Delta X \Delta s\|_{-\infty} \leq \|r\|^2/4 \leq n\mu^k/4. \tag{14}$$

Similarly, if (12) holds at the kth iteration of Algorithm 1, we have

$$\|\Delta X \Delta s\|_{-\infty} \leq \alpha \log(n)\mu^k. \tag{15}$$

From these inequalities and Lemma 9, we can prove Theorem 8 in the same way as we proved Theorem 5. □

6 A potential reduction algorithm

In section 2, we described Algorithm 1 which is a path-following algorithm. Here we propose a potential reduction algorithm which is almost the same as Algorithm 1, but the step size $\bar{\theta}$ is determined as the minimal point of the potential function (1). We show that if we use the neighborhood $\mathcal{N}_\infty^-(\beta)$ for $\beta \in (0,1)$, the algorithm requires the same order of number of iterations as Algorithm 1.

The primal-dual potential function (1) can be written as

$$\psi(x,s) = (\rho - n)\log(x^T s) - \sum_{j=1}^{n} \log\left(\frac{x_j s_j}{x^T s/n}\right) + n\log n$$

The inequality of the geometric and arithmetic means yields

$$-\sum_{j=1}^{n} \log\left(\frac{x_j s_j}{x^T s/n}\right) \geq 0.$$

Hence,

$$(\rho - n)\log(x^T s) + n\log n \leq \psi(x,s).$$

This inequality tells us the exact amount, $-(\rho - n)L + n\log n$, to which ψ should be reduced in order to achieve

$$x^T s \leq 2^{-L}.$$

We assume that we have initial points x^0 of (P) and (y^0, s^0) of (D) such that $(x^0, s^0) \in \mathcal{N}$ and $\psi(x^0, s^0) \leq O((\rho - n)L) + n \log n$.

Now we propose the following potential reduction algorithm based on $\mathcal{N}$, γ, and $\psi(x, s)$.

Algorithm 2 *Change the computational method of $\bar{\theta}$ at each iteration in Algorithm 1 as follows*

$$\psi(x(\bar{\theta}), s(\bar{\theta})) \leq \psi(x(\theta), s(\theta)) \text{ for each } (x(\theta), s(\theta)) \in \mathcal{N}.$$

Then we have the following result.

Theorem 11 *Let $\beta \in (0,1)$ and $\gamma \in (0,1)$ be constants such that $\gamma \leq 2(1-\beta)$. Let $\alpha = 3$, $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$, and*

$$\rho = n + \frac{\alpha n^2}{\beta\gamma(1-\gamma)} \log \frac{1}{1-\beta}.$$

Then Algorithm 2 requires an anticipated number of iterations that is $O(\log(n)L)$, with a worst-case bound of $O(nL)$.

Proof. Let θ_2 be the parameter size defined in Lemma 9. Then we see

$$\psi(x(\bar{\theta}), s(\bar{\theta})) \leq \psi(x(\theta_2), s(\theta_2)).$$

Now we have

$$\begin{aligned}
&\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x^k, s^k) \\
&\leq \psi(x(\theta_2), s(\theta_2)) - \psi(x^k, s^k) \\
&= (\rho - n) \log \frac{x(\theta_2)^T s(\theta_2)}{(x^k)^T s^k} - \sum_{j=1}^{n} \log \frac{x(\theta_2)_j s(\theta_2)_j}{x(\theta_2)^T s(\theta_2)} + \sum_{j=1}^{n} \log \frac{x_j^k s_j^k}{(x^k)^T s^k} \\
&= (\rho - n) \log \frac{\mu(\theta_2)}{\mu^k} - \sum_{j=1}^{n} \log \frac{x(\theta_2)_j s(\theta_2)_j}{\mu(\theta_2)} + \sum_{j=1}^{n} \log \frac{x_j^k s_j^k}{\mu^k} \\
&\leq (\rho - n) \log(1 - (1-\gamma)\theta_2) - \sum_{j=1}^{n} \log(1-\beta) \\
&\quad \text{(from (4) and } (x(\theta_2), s(\theta_2)) \in \mathcal{N}_\infty^-(\beta) \text{)} \\
&\leq -\frac{\alpha n^2 \theta_2}{\beta\gamma} \log \frac{1}{1-\beta} - n \log(1-\beta).
\end{aligned}$$

If $\theta_2 = 1$ then

$$\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x^k, s^k) \leq -\left(\frac{\alpha n^2}{\beta\gamma} - n\right) \log \frac{1}{1-\beta}.$$

Now suppose that $\theta_2 < 1$. From Lemma 9 and (14) we have

$$\theta_2 \geq \frac{4\beta\gamma}{n}.$$

If (12) holds, from (15) we have

$$\theta_2 \geq \frac{\beta\gamma}{\alpha \log(n)}.$$

Thus

$$\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x^k, s^k) \leq -(4\alpha - 1)n \log \frac{1}{1-\beta},$$

and, if (12) holds, then

$$\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x^k, s^k) \leq -\left(\frac{n^2}{\log(n)} - n\right) \log \frac{1}{1-\beta}.$$

Hence we easily obtain the desired results. □

The interesting point is that the choice of ρ here is significantly larger than the theoretically best choice of ρ. This choice is widely used in practical implementations (McShane et al. [10] and Lustig et al. [9]).

References

[1] K. M. Anstreicher and R. A. Bosch, "Long steps in a $O(n^3L)$ algorithm for linear programming," Yale School of Management, (New Haven, CT, 1989).

[2] D. A. Bayer and J. C. Lagarias, "The nonlinear geometry of linear programming: I. Affine and projective scaling trajectories, II. Legendre transform coordinates and central trajectories, *Transactions of the American Mathematical Society* 314 (1989) 499-581.

[3] R. M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," OR 182-88, Sloan

School of Management, Massachusetts Institute of Technology (Cambridge, MA, 1988).

[4] C. C. Gonzaga, "An algorithm for solving linear programming programs in $O(n^3L)$ operations," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1988) 1-28.

[5] C. C. Gonzaga and M. J. Todd, "An $O(\sqrt{n}L)$-iteration large-step primal-dual affine algorithm for linear programming," Technical Report No. 862, School of Operations Research and Industrial Engineering, Cornell University, (Ithaca, New York, 1989).

[6] M. Kojima, S. Mizuno, and A. Yoshise, "An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems," Research Reports on Information Sciences B-217, Dept. of Information Sciences, Tokyo Institute of Technology (Meguro-ku, Tokyo, Japan, 1988).

[7] M. Kojima, S. Mizuno, and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1989) 29-47.

[8] M. Kojima, S. Mizuno and A. Yoshise, "A Polynomial-Time Algorithm for a Class of Linear Complementarity Problems," *Mathematical Programming* 44 (1989) 1-26.

[9] I. J. Lustig, D. F. Shanno and J. W. Gregory, "The primal-dual interior point method on the Cray supercomputer," Department of Civil Engineering and Operations Research, Princeton University (Princeton, NJ, 1989).

[10] K. A. McShane, C. L. Monma, and D. Shanno, "An implementation of a primal-dual interior point method for linear programming" *ORSA Journal on Computing* 1 (1989) 70-83.

[11] N. Megiddo, "Pathways to the optimal set in linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1988) 131-158.

[12] S. Mizuno, M. J. Todd, and Y. Ye, "Anticipated behavior of path-following algorithms for linear programming," Technical Report No. 878, School of Operations Research and Industrial Engineering, Cornell University, (Ithaca, New York, 1989).

[13] R. C. Monteiro, I. Adler, and M. G. C. Resende, "A polynomial time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension," Report ESRC 88-8, Department of Industrial Engineering and operations Research, University of California, (Berkeley, California, 1988).

[14] R. C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part I: Linear programming," *Mathematical Programming* 44 (1989) 27-42.

[15] R. C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part II: Convex quadratic programming," *Mathematical Programming* 44 (1989) 43-66.

[16] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* 40 (1988) 59-94.

[17] S. I. Resnick, "*Extreme values, regular variation, and point processes*," (Springer-Verlag, New York, 1987).

[18] G. Sonnevend, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," in: *Lecture Notes in Control and Information Sciences* 84 (Springer, New York, 1985) 866-876.

[19] M. J. Todd and Y. Ye, "A centered projective algorithm for linear programming," Technical Report No. 763, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, New York, 1987) to appear in *Mathematics of Operations Research*.

[20] P. M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations," AT&T Bell Laboratories (Murray Hill, NJ, 1987) to appear in *Mathematical Programming*.

[21] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," Technical Report, Department of Management Sciences, The University of Iowa (Iowa City, IA, 1989) to appear in *Mathematical Programming*.